TAMS26 – Lecture 9
– Canonical correlation analysis –

Martin Singull

Mathematical Statistics
Department of Mathematics
Linköping University

October 8 – 2012

Content - Lecture

- Canonical correlation analysis
- Sample canonical correlation analysis
- Some test - Asymptotic distribution
- Testing independence
  - Known dependency structure
  - Unknown dependency structure
- Some test - Asymptotic distribution
- Example – fMRI and CCA

Model

Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two normal p_1- and p_2-variate random vectors,

$$\boldsymbol{x}_1 \sim N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$

and

$$\boldsymbol{x}_2 \sim N_{p_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

with expectations $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and covariance matrices $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ for some p_1 and p_2.

Model, cont.

Combine the two vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ into a new vector

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}$$

and let the covariance matrix for the new vector be

$$Cov\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Canonical correlation analysis (CCA) finds the pairs of linear combinations $\boldsymbol{\alpha}'\boldsymbol{x}_1$ and $\boldsymbol{\beta}'\boldsymbol{x}_2$, where $\boldsymbol{\alpha} \in \mathbb{R}^{p_1}$ and $\boldsymbol{\beta} \in \mathbb{R}^{p_2}$, which have the maximum correlation.

$$\begin{aligned}\rho &= max_{\boldsymbol{\alpha},\boldsymbol{\beta}}\, \mathrm{Corr}(\boldsymbol{\alpha}'\boldsymbol{x}_1, \boldsymbol{\beta}'\boldsymbol{x}_2) \\ &= max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}}{(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta})^{1/2}}\end{aligned}$$

The canonical correlation is invariant to scaling of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

$$\begin{aligned}\rho = max_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Big\{\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}\ :\ &\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}, \\ &\boldsymbol{\alpha} \in \mathbb{R}^{p_1},\ \boldsymbol{\beta} \in \mathbb{R}^{p_2}\Big\}\end{aligned}$$

The canonical correlation coefficient is also non-negative, as it otherwise is possible to flip the sign of either $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$.

The first canonical variables u_1 and v_1 are the two linear functions $u_1 = \boldsymbol{\alpha}_1'\boldsymbol{x}_1$ and $v_1 = \boldsymbol{\beta}_1'\boldsymbol{x}_2$ having maximum correlation subject to the condition that $Var(u_1) = Var(v_1) = 1$.

The second canonical variables u_2 and v_2 are the two linear functions $u_2 = \boldsymbol{\alpha}_2'\boldsymbol{x}_1$ and $v_2 = \boldsymbol{\beta}_2'\boldsymbol{x}_2$ having maximum correlation subject to the condition that u_2 and v_2 are uncorrelated with both u_1 and v_1 and have unit variance, and so on.

The canonical correlation coefficients can be found as the solution of one of the following two matrix equations

$$\begin{aligned}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha} &= \rho^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\beta} &= \rho^2\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}\end{aligned}$$

Sample CCA

Assume now, that we have n independent observations and let for convenience $p_1 < p_2$. Then we have two random matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, such that

$$\boldsymbol{X}_1 = (\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, ..., \boldsymbol{x}_{1n}), \text{ where } \boldsymbol{x}_{1i} \sim N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$

and

$$\boldsymbol{X}_2 = (\boldsymbol{x}_{21}, \boldsymbol{x}_{22}, ..., \boldsymbol{x}_{2n}), \text{ where } \boldsymbol{x}_{2i} \sim N_{p_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}).$$

Sample CCA, cont.

Let $1 > r_1^2 > \cdots > r_{p_1}^2 > 0$ be the sample canonical correlation coefficients calculated from $\widehat{\boldsymbol{\Sigma}}_{MLE}$, the maximum likelihood estimate of the covariance matrix $\boldsymbol{\Sigma}$.

The distribution of $r_1^2, r_2^2, ..., r_{p_1}^2$ when $\rho_1^2 = \rho_2^2 = ... = \rho_{p_1}^2 = 0$ is the distribution of the nonzero roots of

$$|\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} - \nu\widehat{\boldsymbol{\Sigma}}_{22}| = 0.$$

Null distribution

The null distribution of $r_1^2, ..., r_{p_1}^2$, i.e., the distribution when $\rho_1^2 = ... = \rho_{p_1}^2 = 0$, is given by Muirhead (1982)

$$f_{r_1^2,...,r_{p_1}^2}(\nu_1, ..., \nu_{p_1}) = C\prod_{i=1}^{p_1}\left\{\nu_i^{\frac{p_2-p_1-1}{2}}(1-\nu_i)^{\frac{n-p_1-p_2-2}{2}}\right\}\prod_{i<j}^{p_1}(\nu_i-\nu_j),$$

where

$$C = \pi^{{p_1}^2/2}\frac{\Gamma_{p_1}(\frac{n-1}{2})}{\Gamma_{p_1}(\frac{n-1-p_2}{2})\Gamma_{p_1+p_2}(\frac{p_1}{2})\Gamma_{p_1+p_2}(\frac{p_2}{2})}$$

and $\Gamma_{p_1}(\cdot)$ is the multivariate Gamma function

$$\Gamma_{p_1}(t) = \pi^{\frac{p_1(p_1-1)}{4}}\prod_{i=1}^{p_1}\Gamma\left(t - \frac{i-1}{2}\right).$$

Some test

We may be interested to test whether the last $p-k$ correlations are zero, for in this case, the last $p-k$ canonical variables $(\boldsymbol{\alpha}_i'\boldsymbol{x}_1, \boldsymbol{\beta}_i'\boldsymbol{x}_2)$, $i = k+1, ..., p$ are of no predictive value, and the relationship between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can be measured by the first k canonical variables.

When there are many variables in each set, this can often lead to a substantial reduction in dimensionality. Formally we wish to test the hypothesis

$$H : \rho_{k+1} = ... = \rho_p = 0, \quad \rho_k > 0,$$

vs.

$$A : \rho_i \neq 0 \text{ at least one } i = k+1, ..., p.$$

The likelihood ratio test is based on the statistic

$$\Lambda_{p-k} = \prod_{i=k+1}^{p} (1 - r_i^2)$$

and one can show that (Fujikoshi, 1977) as $n \to \infty$

$$-\left(n - k - \frac{1}{2}(p_1 + p_2 + 3) - \sum_{i=1}^{k} r_i^2\right) \ln \Lambda_{p-k}$$

is asymptotically distributed as χ^2 with $g = (p_1 - k)(p_2 - k)$ degrees of freedom.

Linköping University
INSTITUTE OF TECHNOLOGY

Let the observation matrix be

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{pmatrix},$$

where $\boldsymbol{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, n$ and $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are independent for $i \neq j$.

The distribution of the observation matrix is written as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{pmatrix} \sim N_{p,n}\left(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I}\right),$$

which is equivalent to

$$\mathrm{vec}\boldsymbol{X} \sim N_{pn}\left(\mathrm{vec}\boldsymbol{M}, \boldsymbol{I} \otimes \boldsymbol{\Sigma}\right).$$

Linköping University
INSTITUTE OF TECHNOLOGY

Let the matrix $\boldsymbol{X} : (p \times n)$ have matrix normal distribution with a separable covariance matrix, i.e.,

$$\boldsymbol{X} \sim N_{p,n}\left(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\right),$$

which is equivalent to

$$\mathrm{vec}\boldsymbol{X} \sim N_{pn}\left(\mathrm{vec}\boldsymbol{M}, \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}\right).$$

The matrix $\boldsymbol{X}$ has the density function

$$f(\boldsymbol{X}) = (2\pi)^{-\frac{1}{2}pn} |\boldsymbol{\Sigma}|^{-n/2} |\boldsymbol{\Psi}|^{-p/2} \\ etr\left\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X} - \boldsymbol{M}\right)\boldsymbol{\Psi}^{-1}\left(\boldsymbol{X} - \boldsymbol{M}\right)'\right\}.$$

Linköping University
INSTITUTE OF TECHNOLOGY

- Known dependency structure, $\boldsymbol{\Psi}$
- Unknown dependency structure, $\boldsymbol{\Psi}$

Linköping University
INSTITUTE OF TECHNOLOGY

We will call $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ the spatial and temporal covariance matrix, respectively.

Furthermore, we assume first that $\boldsymbol{\Psi}$ is known.

Let us assume that

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{pmatrix},$$

where

$$\boldsymbol{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Hence, we have n dependent vectors.

$$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}\mathbf{1}' = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \mathbf{1}' = \boldsymbol{M} : (p \times n).$$

The MLEs are given by

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= (\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{1})^{-1}\boldsymbol{X}\boldsymbol{\Psi}^{-1} \\
n\widehat{\boldsymbol{\Sigma}} &= \left(\boldsymbol{X} - \widehat{\boldsymbol{\mu}}\mathbf{1}'\right)\boldsymbol{\Psi}^{-1}\left(\boldsymbol{X} - \widehat{\boldsymbol{\mu}}\mathbf{1}'\right)' \\
&= \boldsymbol{X}\left(\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\mathbf{1}\left(\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\boldsymbol{\Psi}^{-1}\right)\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}' = \boldsymbol{A}.
\end{aligned}$$

Special case: If $\boldsymbol{\Psi} = \boldsymbol{I}_n$ we have

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= (\mathbf{1}'\mathbf{1})^{-1}\boldsymbol{X}\mathbf{1} = \frac{1}{n}\boldsymbol{X}\mathbf{1}, \\
n\widehat{\boldsymbol{\Sigma}} &= \boldsymbol{X}\left(\boldsymbol{I}_n - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{C}\boldsymbol{X}',
\end{aligned}$$

where $\boldsymbol{C} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$.

Let us partition $\boldsymbol{X}$, $\boldsymbol{M}$ and $\boldsymbol{\Sigma}$ into k parts, as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_k \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_k \end{pmatrix} \quad \text{and}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & & \boldsymbol{\Sigma}_{2k} \\ \vdots & \vdots & \ddots & \\ \boldsymbol{\Sigma}_{k1} & \boldsymbol{\Sigma}_{k2} & & \boldsymbol{\Sigma}_{kk} \end{pmatrix},$$

where $\boldsymbol{X}_i : (p_i \times n)$, $\boldsymbol{\mu}_i : (p_i \times 1)$ and $\boldsymbol{\Sigma}_{ij} : (p_i \times p_j)$ for $i, j = 1, \ldots, k$.

We wish to test spatial independence,

$$H : \boldsymbol{\Sigma}_{ij} = 0, \text{ when } i \neq j.$$

Define $\boldsymbol{\Sigma}^*$ to be the covariance matrix between the rows when H is true

$$\boldsymbol{\Sigma}^* \equiv \begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_{22} & & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \boldsymbol{\Sigma}_{kk} \end{pmatrix}.$$

Theorem

The likelihood ratio test of level α for testing the null hypothesis H rejects H if $\lambda \leqslant c_\alpha$, where

$$\lambda = \frac{|\boldsymbol{A}|}{\prod_{i=1}^{k} |\boldsymbol{A}_{ii}|},$$

$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}'$, $\boldsymbol{A}_{ii} = \boldsymbol{X}_i\boldsymbol{H}\boldsymbol{X}_i'$ and $\boldsymbol{H}$ is the weighted centralization matrix

$$\boldsymbol{H} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\mathbf{1}\left(\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\boldsymbol{\Psi}^{-1}.$$

Choose c_α such that the significance level of the test is α.

Asymptotic distribution

For large n, $-m \ln \lambda$ is asymptotically distributed as χ^2 with g degrees of freedom, where

$$m = n - 2a,$$

$$a = \frac{p^3 - \sum_{i=1}^{k} p_i^3 + 9\left(p^2 - \sum_{i=1}^{k} p_i^2\right)}{6\left(p^2 - \sum_{i=1}^{k} p_i^2\right)},$$

$$g = \sum_{j=2}^{k}\sum_{i=1}^{j-1} p_i p_j = \frac{1}{2}\left(p^2 - \sum_{i=1}^{k} p_i^2\right).$$

Furthermore, we have

Theorem

Let $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, where $\boldsymbol{M} = \boldsymbol{\mu}\mathbf{1}'$ and $\boldsymbol{\Psi}$ is known. Then $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}' \sim W_p(n-1, \boldsymbol{\Sigma})$.

Proof.

$\boldsymbol{A}$ is Wishart since $\boldsymbol{H}\boldsymbol{\Psi}$ is idempotent. Hence, $\boldsymbol{A} \sim W_p(\text{rank}(\boldsymbol{H}), \boldsymbol{\Sigma}, \boldsymbol{M}\boldsymbol{H}\boldsymbol{M}')$, but $\text{rank}(\boldsymbol{H}) = n-1$ and $\boldsymbol{M}\boldsymbol{H}\boldsymbol{M}' = \mathbf{0}$ so

$$\boldsymbol{A} \sim W_p(n-1, \boldsymbol{\Sigma}).$$

□

Since $\boldsymbol{A} \sim W_p(n-1, \boldsymbol{\Sigma})$ several properties follows, such as

- invariance,
- asymptotic null distribution.

See Muirhead (1982) for more details.

Invariance $(k = 2)$

$$\lambda = \frac{|\boldsymbol{A}|}{|\boldsymbol{A}_{11}||\boldsymbol{A}_{22}|} = \prod_{i=1}^{p_1}(1 - r_i^2)$$

where $r_1^2, \ldots, r_{p_1}^2$ are the ch. roots of $\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$.

The distribution of λ depends only on $\rho_1^2, \ldots, \rho_{p_1}^2$ the ch. roots of $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ (the canonical correlation coefficients).

$$H : \rho_1 = \ldots = \rho_{p_1} = 0$$

Asymptotic Null Distribution $(k = 2)$

One can show that the test statistic

$$-m \ln \lambda = -\left(n - \frac{1}{2}(p_1 + p_2 + 3)\right) \sum_{i=1}^{p_1} \ln(1 - r_i^2),$$

asymptotically follows a χ_g^2 distribution with $g = p_1 p_2$.

For p-values one can use the formula

$$\begin{aligned} \mathrm{P}(-m \ln \lambda \geq z) = &\mathrm{P}(\chi_g^2 \geq z) + \\ &+ \frac{1}{m^2}\gamma^2(\mathrm{P}(\chi_{g+4}^2 \geq z) - \mathrm{P}(\chi_g^2 \geq z)) + O(m^{-3}), \end{aligned}$$

where $\gamma^2 = \frac{1}{48}p_1p_2(p_1^2 + p_2^2 - 5)$.

Unknown Dependency Structure

Assume that $\boldsymbol{\Psi}$ is unknown, but follows some structure.

Define

$$\widehat{\boldsymbol{A}} \equiv \boldsymbol{X}\widehat{\boldsymbol{H}}\boldsymbol{X}'$$

and

$$\widehat{\boldsymbol{A}}_{ii} \equiv \boldsymbol{X}_i\widehat{\boldsymbol{H}}\boldsymbol{X}_i',$$

where $\widehat{\boldsymbol{H}}$ is the estimated weighted centralization matrix i.e.,

$$\widehat{\boldsymbol{H}} \equiv \widehat{\boldsymbol{\Psi}}^{-1} - \widehat{\boldsymbol{\Psi}}^{-1}\mathbf{1}\left(\mathbf{1}'\widehat{\boldsymbol{\Psi}}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\widehat{\boldsymbol{\Psi}}^{-1}.$$

Theorem

Suppose that the temporal covariance matrix $\mathbf{\Psi}$ can be estimated explicitly. The likelihood ratio test of level α for testing the null hypothesis H rejects H if $\Lambda \leqslant c_\alpha$, where the statistic is

$$\Lambda = \left(\frac{|\widehat{\mathbf{\Psi}}|}{|\widehat{\mathbf{\Psi}}_0|}\right)^{p/2}\left(\frac{|\widehat{\boldsymbol{A}}|}{\prod_{i=1}^{k}|\widehat{\boldsymbol{A}}_{ii}|}\right)^{n/2}.$$

Choose c_α such that the significance level of the test is α.

Let $\mathbf{\Psi}$ be the covariance matrix from an autoregressive process of order one, AR(1)

$$\mathbf{\Psi}(\theta) = \frac{1}{1-\theta^2}\begin{pmatrix} 1 & \theta & \theta^2 & \cdots & \theta^{n-1} \\ \theta & 1 & \cdots & & \\ \theta^2 & \vdots & \ddots & & \\ \vdots & & & & \theta \\ \theta^{n-1} & & & \theta & 1 \end{pmatrix}$$

and let the covariance matrix $\mathbf{\Sigma}$ be unstructured.

$$x_{it} - \mu_i = \theta\,(x_{it-1} - \mu_i) + \varepsilon_{it}, \quad i = 1, \ldots, p, t = 1, \ldots, n$$

for some expectations μ_i and where $\varepsilon_{it} \sim N(0, \mathbf{\Sigma}_{ii}), |\theta| < 1$ and ε_{it} is uncorrelated with x_{is} for each $s < t$.

The different AR(1) time series are dependent since $\mathbf{\Sigma}_{ij} \neq 0$.

The determinant and the inverse of $\mathbf{\Psi}(\theta)$ can easily be calculated,

$$|\mathbf{\Psi}(\theta)| = (1-\theta^2)^{-1}$$

and

$$\mathbf{\Psi}^{-1}(\theta) = \boldsymbol{I} + \theta^2\boldsymbol{D}_1 - \theta\boldsymbol{D}_2,$$

where $\boldsymbol{D}_1 = diag(0, 1, \ldots, 1, 0)$ and $\boldsymbol{D}_2$ is tridiagonal matrix with zeros on the diagonal and ones on the super- and subdiagonal.

Likelihood function

The likelihood function can be written as

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \theta) = |\boldsymbol{\Sigma}|^{-n/2}(1-\theta^2)^{p/2} \mathrm{etr}\left\{-\frac{1}{2}\left(\boldsymbol{X}-\boldsymbol{\mu}\mathbf{1}'\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}-\boldsymbol{\mu}\mathbf{1}'\right)\boldsymbol{\Psi}^{-1}(\theta)\right\}.$$

Likelihood equations

We have the likelihood equations

$$\boldsymbol{\mu} = (\mathbf{1}'\boldsymbol{\Psi}^{-1}(\theta)\mathbf{1})^{-1}\boldsymbol{X}\boldsymbol{\Psi}^{-1}(\theta)\mathbf{1},$$
$$n\boldsymbol{\Sigma} = \boldsymbol{X}\boldsymbol{H}(\theta)\boldsymbol{X}'$$

and

$$2\theta p + (1-\theta^2)\mathrm{tr}\left\{\Upsilon(\theta)\left(\boldsymbol{X}-\boldsymbol{\mu}\mathbf{1}'\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}-\boldsymbol{\mu}\mathbf{1}'\right)\right\} = 0,$$

where

$$\Upsilon(\theta) = 2\theta\boldsymbol{D}_1 - \boldsymbol{D}_2.$$

Insertion and we have the equation

$$2\theta p + n(1-\theta^2)\mathrm{tr}\left\{\boldsymbol{X}\boldsymbol{B}\Upsilon(\theta)\boldsymbol{B}'\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{B}\boldsymbol{\Psi}^{-1}(\theta)\boldsymbol{B}'\boldsymbol{X}')^{-1}\right\} = 0,$$

where

$$\boldsymbol{B} = \boldsymbol{I} - (\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{1})^{-1}\boldsymbol{\Psi}^{-1}\mathbf{1}\mathbf{1}'.$$

Polynomial equation of order less than $3p+2$ $(2p+3)$.

Algorithm for the MLE

1. Obtain an initial estimate $\widehat{\theta}^{(0)}$ of θ, using the mean of the Yule-Walker estimates.
2. Compute $\widehat{\boldsymbol{\mu}}^{(1)}$ and $\widehat{\boldsymbol{\Sigma}}^{(1)}$ from the likelihood equations for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, using $\widehat{\theta}^{(0)}$.
3. Compute the value of $\widehat{\theta}^{(k)}$ by solving the likelihood equation for θ using the estimates $\widehat{\boldsymbol{\mu}}^{(k)}$ and $\widehat{\boldsymbol{\Sigma}}^{(k)}$. Ensure that $|\widehat{\theta}^{(k)}| < 1$ and that the solution is a maximum.
4. Compute $\widehat{\boldsymbol{\mu}}^{(k)}$ and $\widehat{\boldsymbol{\Sigma}}^{(k)}$ from the likelihood equations for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, using the estimate $\widehat{\theta}^{(k)}$ from previous step.
5. Repeat Steps 3 and 4 until convergence is obtained, i.e., until

$$|\widehat{\theta}^{(k)} - \widehat{\theta}^{(k-1)}| < \varepsilon \quad \text{and} \quad \mathrm{tr}\left((\widehat{\boldsymbol{\Sigma}}^{(k)} - \widehat{\boldsymbol{\Sigma}}^{(k-1)})^2\right) < \varepsilon.$$

Under the null hypothesis we have the likelihood equations

$$\begin{aligned}\boldsymbol{\mu}_i &= (\mathbf{1}'\boldsymbol{\Psi}^{-1}(\theta)\mathbf{1})^{-1}\boldsymbol{X}_i\boldsymbol{\Psi}^{-1}(\theta)\mathbf{1},\\ n\boldsymbol{\Sigma}_{ii} &= \boldsymbol{X}_i\underline{\mathrm{H}}(\theta)\boldsymbol{X}_i'\end{aligned}$$

for $i = 1, \ldots, k$ and

$$2\theta p + (1-\theta^2) \quad \operatorname{tr}\left\{\Upsilon(\theta)\left(\sum_{i=1}^{k}(\boldsymbol{X}_i - \boldsymbol{\mu}_i\mathbf{1}')'\,\boldsymbol{\Sigma}_{ii}^{-1}\,(\boldsymbol{X}_i - \boldsymbol{\mu}_i\mathbf{1}')\right)\right\} = 0.$$

Linköping University
INSTITUTE OF TECHNOLOGY

1. Obtain an initial estimate $\widehat{\theta}_0^{(0)}$, of θ as the mean of the Yule-Walker estimates.
2. Compute $\widehat{\boldsymbol{\mu}}_i^{(1)}$ and $\widehat{\boldsymbol{\Sigma}}_{ii}^{(1)}$, for $i = 1, \ldots, k$, from the likelihood equations for $\boldsymbol{\mu}_i$ and Σ_{ii}, using $\widehat{\theta}_0^{(0)}$.
3. Compute the value of $\widehat{\theta}_0^{(l)}$ by solving the likelihood equation for θ using the estimates $\widehat{\boldsymbol{\mu}}_i^{(l)}$ and $\widehat{\boldsymbol{\Sigma}}_{ii}^{(l)}$. Ensure that $|\widehat{\theta}_0^{(l)}| < 1$ and that the solution is a maximum.
4. Compute $\widehat{\boldsymbol{\mu}}_i^{(l)}$ and $\widehat{\boldsymbol{\Sigma}}_{ii}^{(l)}$ from the likelihood equations for $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_{ii}$, using the estimate $\widehat{\theta}_0^{(l)}$ from previous step.
5. Repeat Steps 3 and 4 until convergence is obtained,

$$|\widehat{\theta}_0^{(k)} - \widehat{\theta}_0^{(k-1)}| < \varepsilon \quad \text{and} \quad \operatorname{tr}\left((\widehat{\boldsymbol{\Sigma}}_{ii}^{(k)} - \widehat{\boldsymbol{\Sigma}}_{ii}^{(k-1)})^2\right) < \varepsilon, \forall i.$$

ch.

Linköping University
INSTITUTE OF TECHNOLOGY

Consequently, the likelihood ratio statistic is

$$\Lambda = \left(\frac{1-\widehat{\theta}_0^2}{1-\widehat{\theta}^2}\right)^{p/2}\left(\frac{|\widehat{\mathrm{A}}|}{\prod_{i=1}^{k}|\widehat{\mathrm{A}}_{ii}|}\right)^{n/2}.$$

See Ohlson (2009) for more details.

Linköping University
INSTITUTE OF TECHNOLOGY

How to use CCA in fMRI study?

"Even though it is the temporal behaviour of the voxel time series that determines whether voxels are active or not, the voxel time series exist in a spatial context that can be exploited to improve the detection of active voxels." (Friman, 2003)

Linköping University
INSTITUTE OF TECHNOLOGY

- Temporal basis functions ⇒

  ”Describe what we are searching for in the fMRI data.”

- Spatial basis functions ⇒

  ”Describe what we have in the fMRI data.”

Linköping University
INSTITUTE OF TECHNOLOGY

- The Fourier model ⇒

  “Sine and cosine functions with different frequencies.”

- The Taylor subspace model ⇒

  ”Taylor expansion of any continuous model of the BOLD response. Use the different derivatives as temporal basis functions.”

- PCA generated subspace models ⇒

  ”Choose which derivatives in the Taylor subspace model to use as temporal basis functions.”

Linköping University
INSTITUTE OF TECHNOLOGY

- The single voxel model ⇒

  ”Active brain areas consist of isolated single voxels”

- The Gaussian model ⇒

  ”Gaussian filter, which should be matched to the size of active brain area.”

- The orientation adaptive model ⇒

  ”Several filters, isotropic and anisotropic.”

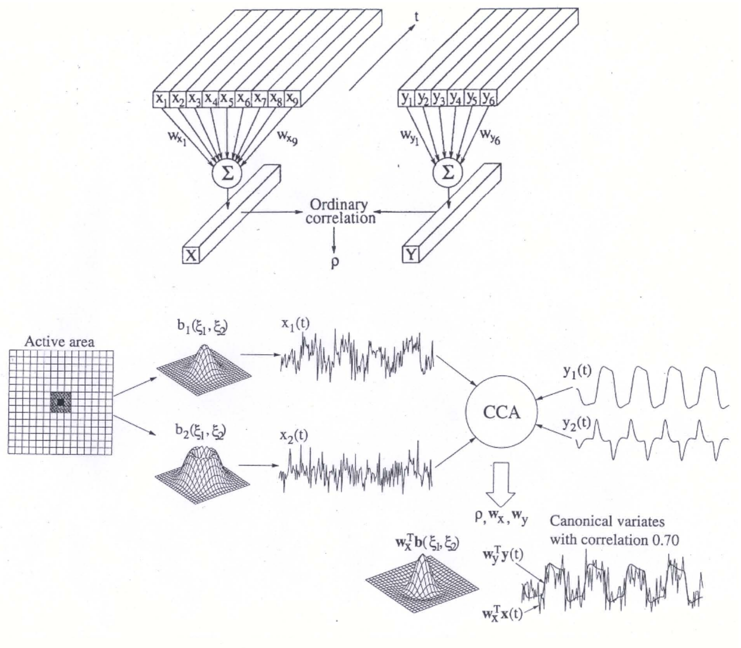
Linköping University
INSTITUTE OF TECHNOLOGY

In fMRI analysis, we want to decide whether a voxel is active or not. The first (largest) canonical correlation coefficient is calculated between the set of filtered real BOLD signals (spatial basis functions) and the set of temporal basis functions (Friman, 2003).

If the canonical correlation coefficient is large enough, it implies a high degree of similarity, and we have an activated voxel.

Linköping University
INSTITUTE OF TECHNOLOGY

Example, cont.



Example, cont.

At each voxel we would like to test the hypothesis,

$$H: \ \rho_1 = 0$$

vs.

$$A: \ \rho_1 > 0$$

and we reject H if the first canonical correlation coefficient, ρ_1, is greater than some threshold, ρ_t.

Example, cont.

fMRI data is dependent data, hence the density function for the canonical correlation coefficients is not valid.

Instead of test $H_0: \ \rho_1 = 0$, Nandy (2003) suggest that it is better to test all the canonical correlation coefficients, i.e.,

$$H_0: \ \rho_1 = \cdots = \rho_{p_1} = 0.$$

This test is equivalent to test independence between two sets of variables.